Computational Network Biology
Biostatistics & Medical Informatics 826
Fall 2018

Sushmita Roy
sroy@biostat.wisc.edu
https://compnetbiocourse.discovery.wisc.edu

Sep 6th 2018
Goals for today

• Administrivia
• Course topics
• Short survey of interests/background
BMI826 Computational Network Biology

• Course home page: https://compnetbiocourse.discovery.wisc.edu
• Instructor: Prof. Sushmita Roy
  – sroy@biostat.wisc.edu
• Office: room 3168, Wisconsin Institute for Discovery
  • Your WISC cards will be enabled for upper floor access.
• Office hours:
  – Tuesday/Thursday: 2:30pm-3:30 pm
  – By appointment via email
• Class announcements via piazza
Finding my office WID 3168
Course organization

• Tentative schedule: https://compnetbiocourse.discovery.wisc.edu/schedule-2/
• The material in this course is organized into five major topics
• At the beginning of each topic I will provide an introduction for the topic
• Most of the material is from published papers and review articles
  – We will read and discuss papers from each of these topics
    • Please see syllabus at https://compnetbiocourse.discovery.wisc.edu/syllabus/
  – Readings will be made online on the schedule page.
• Last week or so will be project presentations
Recommended background

• Computer science
  – Introductory course in data structures is good, but not required

• Statistics
  – Good if you’ve had at least one course, but not required

• Molecular biology
  – Good if you have had some introductory course
  – An interest in learning some basic molecular biology

• Programming background
  – Familiarity with a Linux environment
  – Be able to run programs on data on the command line
  – Be able to write code to do some data analysis and computations
Course grading

- **Written critiques: 20%**
  - Five written critiques: one for each major topic

- **Written and implementation assignments: 30%**
  - Three or so

- **Project: 45%**
  - Proposal 10%
  - Report 20%
  - In class presentation 15%

- **In class participation: 5%**
Writing a critique

• Critiques are due at the end of each major topic
• Critiques will be a 1-2 page analysis of the papers read in each topic
  – Specific papers to be included in the critique will be mentioned
• The critique should have the following components
  – Overview of the problem area
  – Approaches discussed
  – Strengths and weaknesses
  – Extensions to any of the approaches
Project information

• There are three main components to the project
  – Proposal, In class presentation, Project report
• Project proposal draft (Oct 4\textsuperscript{th})
• Project proposal final draft (Oct 18\textsuperscript{th})
• Project presentations in last week of class
• Project report due (Dec 12\textsuperscript{th})
  – Last day of lecture
Computational resources for this class

- Linux server “mi1.biostat.wisc.edu” available through the BMI department
- Please connect to mi1.biostat.wisc.edu
- Accounts for all registered students in the class have been requested
Goals for today

• Administrivia
• Course topics
• Short survey of background and interests
What is Network biology?

• A collection of algorithms and tools to build, interpret, and use graph representations of interacting molecular entities in biological and bio-medical problems (Adapted from Winterbach et al, 2013, BMC Systems Biology)
  – ~15 years old
  – The term Network biology was likely coined by Albert-László Barabási & Zoltán N. Oltvai 2004
  – Intersects with computer science, statistics, physics, molecular biology

• Related/overlapping areas
  – Bioinformatics, Systems biology, Complex systems, Biological network analysis, Network science
Why Network biology?

• Cells are complex systems
  – A complex system: many components that interact to determine overall function
  – Networks are natural representations of complex systems
• Provides a framework and important tools for integration, interpretation and discovery
• Many biological applications e.g.
  – Understanding complex biological processes at the molecular level
  – Disease prognosis
  – Interpretation of genetic variation
  – Predictive models of cellular function
  – Gene function prediction and prioritization
Overview of lecture topics

Course material is organized by the biological problem and computational approaches to address the problem

<table>
<thead>
<tr>
<th>Biological problem</th>
<th>Computational approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Mapping regulatory network structure</td>
<td>- Probabilistic graphical models</td>
</tr>
<tr>
<td>- Dynamics and context specificity of networks</td>
<td>- Graph structure learning</td>
</tr>
<tr>
<td>- Understanding design principles of biological networks</td>
<td>- Multiple network learning</td>
</tr>
<tr>
<td>- Interpretation of sequence variants</td>
<td>- Graph clustering</td>
</tr>
<tr>
<td>- Identification of important genes</td>
<td>- Graph alignment</td>
</tr>
<tr>
<td>- Predicting the function of a gene</td>
<td>- Diffusion on graphs</td>
</tr>
</tbody>
</table>
Network inference: How do molecular entities interact within a cell?

- Measure component states
- Select candidate regulators and outputs
- Perturb candidate regulators and measure outputs
- Reconstruct network

Select regulators:
- Transcription factors
- Chromatin factors
- IncRNAs
- Kinases
- Signalling proteins

Select outputs:
- mRNA levels
- Chromatin states

Internal cell state

Input or stimulus

Perturb output

Output 1

Output 2

Amit et al., Nat. Rev. Immunology, 2011
Computational concepts
1. Different types of graphical models for network representation
2. Learning graphical models from data
3. Integrating prior information into models

Network inference

Samples

Gene expression levels

Algorithm

$X_1$, $X_5$, $X_2$

$Y_1$, $Y_2$

Biological knowledge bases

3. Integrating prior information into models
2. Learning graphical models from data
1. Different types of graphical models for network representation
Network dynamics: How do networks change between different biological contexts?

Contexts can be different time points, cell types, disease states, organisms.

Computational concepts
1. Multi-task learning
2. Dynamic models for networks

Gene network rewiring during cell cycle
From Curtis et al., BMC Bioinformatics 2012
Network topology: How is a network organized?

**Computational concepts**
1. Degree distribution
2. Network motifs
3. Centrality measures

Graph clustering: functional and disease module identification

Mitra et al., Nat Rev Genetics 2013;

Barabasi et al., Nat Rev Genetics 2011

Computational concepts
1. Graph clustering
2. Modularity measures
Graph alignment: What parts of networks from two species are similar?

1. Clustering on graphs
2. Scoring subnetworks and subnetwork search
3. Matrix factorization
Graph diffusion: Which genes are most important?

Another measure of performance of the algorithm is the receiver-operating characteristic (ROC) analysis, which plots the true-positive rate (TPR) versus the false-positive rate (FPR) subject to the threshold separating the prediction classes. The TPR/FPR is the rate of correctly/incorrectly classified samples of all samples classified to class $+_1$. For evaluating rankings of disease-gene predictions, ROC values can be interpreted as a plot of the frequency of the disease genes above the threshold versus the frequency of disease genes below the threshold, where the threshold is a specific position in the ranking.

In order to compare different curves obtained by ROC analysis, we calculate the area under the ROC curve (AUROC) for each curve.

Results

In this work, we constructed an interaction network based on a total of 35,910 interactions between human proteins as well as 38,975 mapped interactions from four other species. Additionally predicted protein interactions from the STRING database were used (Table 1). We adapted a global distance measure based on random walk with restart (RWR) to define similarity between genes within this interaction network and to rank candidates on the basis of this similarity to known disease genes. Intuitively, the RWR algorithm calculates the similarity between two genes, $i$ and $j$, on the basis of the likelihood that a random walk through the interaction network starting at gene $i$ will finish at gene $j$, whereby all possible paths between the two genes are taken into account. In our implementation, we let the random walk start with equal probability from each of the known disease-gene family members in order to search for an additional family member in the linkage interval (Figure 1). For comparison, we also implemented a similar global search algorithm based on the diffusion kernel (DK), which conceptually performs a different type of random walk calculated by matrix exponentiation (see Material and Methods for mathematical details). In order to compare the performance of global and local network search algorithms, we implemented two previous

Computational concepts

1. Random walks on graphs
2. Graph diffusion kernels
3. Random walks on graphs
Field of view: Graph diffusion: Characterizing genetic variation and impact on complex phenotypes

Computational concepts
1. Graph diffusion methods
2. Subnetwork identification

Graph-based data integration

Computational concepts
1. Graph clustering
2. Clustering multiple graphs

From Wang et al. Nature Methods 2013
Plan for next few lectures

• Sep 11\textsuperscript{th}, 13\textsuperscript{th}
  – Background into graph theory, probability theory and molecular networks
  – Readings:
    • Winterbach et al., Topology of molecular interaction networks. BMC Systems Biology, 2013
      – Section on Network Biology

• Sep 15\textsuperscript{th}
  – Probabilistic graphical models for molecular networks
Learning goals of this class

• Gain a broad overview of the application areas and computational solutions in Network biology
• Gain a deeper understanding one or two areas introduced
• Apply the computational concepts to similar problems in biology and complex systems
• Understand and critique scientific articles
• Enable self learning and deeper study of related topics
Goals for today

- Administrivia
- Course topics
- Short survey of interests/background